**Wine Quality Analysis Documentation**

**Objective:**

The objective of this analysis is to explore the factors affecting the quality of red wine and develop a predictive model to estimate wine quality based on its chemical properties.

**Data Description:**

The dataset used for analysis contains various chemical properties of red wine samples along with their quality ratings. It consists of 1599 observations and 12 features including:

* fixed acidity
* volatile acidity
* citric acid
* residual sugar
* chlorides
* free sulphur dioxide
* total sulphur dioxide
* density
* pH
* sulphates
* alcohol
* quality (target variable)

**Analysis Steps:**

1. **Data Loading and Exploration**: Loaded the dataset from a CSV file and explored its structure using pandas methods like **head()**, **info()**, and **describe()** to understand the data.
2. **Data Visualization**: Visualized the distribution of wine quality ratings using a bar plot and explored pairwise relationships between features using pair plots and correlation matrices.
3. **Feature Engineering**: Created two new features, 'total\_acidity' (sum of fixed acidity, volatile acidity, and citric acid) and 'log\_density' (logarithm of density), to potentially improve model performance.
4. **Model Training and Evaluation**: Split the data into training and testing sets, trained a linear regression model, and evaluated its performance using mean squared error (MSE). Additionally, performed hyperparameter tuning using Ridge regression and cross-validated the models using k-fold cross-validation.
5. **Results Interpretation**: Analyzed the coefficients of the linear regression model to understand the impact of each feature on wine quality.

**Results:**

* Mean Squared Error (MSE) of the linear regression model: 0.3900710908635001
* Best parameters for Ridge regression: {'alpha': 0.1}, Best score: 0.43936491742028033
* Cross-validated MSE using linear regression: 0.4375050167126939
* Coefficients of linear regression model:

| **Feature** | **Coefficient** |
| --- | --- |
| fixed acidity | 0.305345 |
| volatile acidity | -0.722139 |
| citric acid | 0.135807 |
| residual sugar | 0.007582 |
| chlorides | -1.799050 |
| free sulfur dioxide | 0.005651 |
| total sulfur dioxide | -0.003619 |
| density | -1948.894478 |
| pH | -0.390627 |
| sulphates | 0.838648 |
| alcohol | 0.283652 |
| total\_acidity | -0.280988 |

**Conclusion:**

The analysis indicates that several chemical properties of red wine, such as volatile acidity, chlorides, sulphates, and alcohol content, have significant effects on wine quality. The developed models can be used to predict wine quality based on its chemical composition, providing valuable insights for winemakers and quality control processes.